

基于光谱分类的土壤盐分含量预测*

代希君^{1, 2} 彭杰^{3†} 张艳丽³ 罗华平^{1, 2} 向红英³

(1 塔里木大学机械电气化工程学院, 新疆阿拉尔 843300)

(2 塔里木大学现代农业工程重点实验室, 新疆阿拉尔 843300)

(3 塔里木大学植物科学学院, 新疆阿拉尔 843300)

摘要 基于相似土壤组分和光谱特征, 利用土壤光谱反射率数据和曲线特征来进行土壤光谱分类, 同时充分挖掘有效信息是光谱分析的重要应用方向之一。借助模糊 k -均值聚类方法将土壤光谱数据分成四个类别(分类前先将原始光谱进行范围归一化处理), 比较分析了不同类型土壤在光谱分类前后的高光谱特征, 然后利用Kennard-Stone法将各类别样本划分为建模集和预测集, 将预处理后的建模集光谱数据作为输入量, 采用偏最小二乘回归法(PLSR)方法建立全局和各自类别的盐分预测模型。结果表明: 光谱分类建模较按土壤系统分类建模和全局建模的精度有明显提高, 其预测模型总体的预测决定系数 R_p^2 、预测均方根误差RMSEP、相对分析误差RPD和RPIQ(样本观测值三四分位数Q3与一四分位数Q1之差与RMSEP的比值)四个指标分别从0.664、1.219、1.733和1.461提高至0.818、1.132、2.356和2.422, 其中RPD提高幅度达23.13%, 四个类别所建模型RPD均大于2.0, 可以对土壤含盐量进行较为精确的定量研究。研究结果为利用大样本光谱数据建立大尺度区域的盐分等土壤属性预测模型提供一种新的思路和方法。

关键词 土壤; 高光谱; 盐分; 模糊 k -均值聚类; 盐渍化; Kennard-Stone法

中图分类号 TP79; S127 **文献标识码** A

土壤盐渍化是干旱可持续发展和改善环境质量的战略问题之一^[1-2]。我国盐渍土面积大, 分部广, 堪称世界之最^[3]。据统计, 我国耕地中的盐渍化面积达到 $9.2 \times 10^6 \text{ hm}^2$, 占全国耕地面积的6.62%^[4]。新疆是我国盐渍化大区, 盐碱土面积达 $2.18 \times 10^7 \text{ hm}^2$, 现有耕地中31.1%的面积受到盐碱危害, 大面积的盐渍化土壤导致很大一部分农作物因受不同程度盐害而减产绝收, 难以发挥其潜力产量, 土壤盐渍化问题和灌溉引起的土壤次生盐渍化问题严重制约了新疆农业的进一步发展, 也是影响绿洲生态环境稳定的重要因素^[5-7]。因此, 对盐渍化耕地进行治理与改良及种植结构调整与优化, 对新疆乃至我国农业的可持续发展具有深远意义。而这些决策的制定与执行的根本依据是土壤盐渍化

信息。如何快速、准确、低廉地获取大面积盐渍化土壤的动态信息, 并据此合理制定盐渍化土壤治理、改良等方面的决策, 是农业科学研究的新主题^[4]。遥感技术具有获取信息量大、多平台和多分辨率(时间和空间)、快速、覆盖面积大的优势, 是及时掌握大面积盐渍化等土壤信息的最佳手段^[8], 高光谱遥感更是进行土壤盐渍化监测较为理想的一种手段^[9]。盐渍化土壤近地面高光谱试验的一系列研究, 为其高光谱遥感监测的实施和应用提供了一定的理论依据和技术支撑。

国内外研究者在使用高光谱技术获取土壤养分和理化特性信息方面已经取得了一定的成果。Gogé等^[10]在法国本土采集2 200个土样光谱, 借助Local算法进行当地土壤有机碳、CEC、沙土和黏

* 国家自然科学基金项目(41261083, 41361048, 11464039)资助Supported by the National Natural Science Foundation of China (Nos. 41261083, 41361048, 11464039)

† 通讯作者 Corresponding author, E-mail: pjzky@163.com

作者简介: 代希君(1989—), 男, 湖北孝感人, 硕士研究生, 主要从事近地传感器应用研究。E-mail: dxj-sp@163.com

收稿日期: 2015-11-17; 收到修改稿日期: 2016-01-25; 优先数字出版日期(www.cnki.net): 2016-03-1



土等属性的快速估测。Allbed等^[11]利用IKONOS获取阿拉伯半岛Al-Hassa绿洲的高分辨率遥感影像数据,通过对比13种光谱指数反演土壤电导率的精度,发现SAVI, NDSI和SI-T指数在评估以枣椰树为主的地区土壤盐渍化状况效果最好。Weng等^[12]利用ASD光谱仪获得的近地面高光谱数据和Hyperion数据,定量反演了黄河流域土壤盐分含量。屈永华等^[13]以内蒙古河套灌区为试验区,建立了土壤盐分与高光谱数据的偏最小二乘回归(PLSR)模型。独立验证结果表明,对全盐(S%)、硫酸根离子(SO₄²⁻)、pH以及钾、钠总含量(K⁺+Na⁺)有较好的反演精度。王静等^[14]以松嫩平原西部长岭县为例,利用盐渍土高光谱数据构建盐渍土盐分遥感预测模型。结果表明与传统回归相比($R^2=0.81$, RMSE=0.25),运用高光谱数据与人工神经网络方法相结合,能够提高盐渍土的预测精度。刘焕军等^[15]以松嫩平原盐碱土为研究对象,确定了土壤盐碱化程度的最佳波段与土壤光谱指数,建立的高光谱定量模型可以用于评价土壤盐碱化程度。纪文君等^[16]利用高光谱技术分析了不同地区及类型的791个土壤样品,提出了600~800 nm是有机质的响应波段,为大区域建立通用模型快速测定土壤有机质含量提供了可能。王乾龙等^[17]收集13个省的1 661个土样数据,采用PLSR、LWR等方法建立高光谱全氮反演模型,利用浙江省104个水稻土对模型进行了验证,并取得了较好的效果。

本研究基于相似土壤组分和光谱特征,利用土壤光谱反射数据和曲线特征来进行土壤光谱分类,确定最佳分类数目,然后采用PLSR方法构建模型,探求南疆不同类型土壤的盐分光谱预测方法。

1 材料与方法

1.1 研究区概况

阿克苏地区位于新疆维吾尔自治区南疆中部,地处天山南麓,塔里木盆地北缘,地理位置为78°03'~84°07' E, 39°30'~42°41' N,阿克苏河的冲积平原带。年降水量42.4~94.4 mm,蒸发量1 200~1 500 mm。和田地区位于新疆维吾尔自治区最南端,南枕昆仑山与西藏自治区交界,北临塔克拉玛干大沙漠与阿克苏地区的阿瓦提、沙雅接壤,地理位置为77°31'~84°55' E, 34°22'~39°38' N,

年均降水量35 mm,年均蒸发量高达2 480 mm,终年处于极端干旱状态。本研究选取的研究范围涉及温宿县、拜城县、阿瓦提县、新和县、和田县(属和田地区)。该区域农作物以棉花、水稻、玉米、冬小麦为主,还有红枣、核桃、苹果、香梨、杏、葡萄等瓜果,两个地区均属温带大陆性气候,由于蒸发强烈,极易形成盐渍化土壤,阻碍作物生长造成盐害。给当地的农业发展造成极大的困扰,在一定程度上制约了经济的发展。

1.2 样品采集与处理

根据代表性、典型性和集中性原则,供试的339个样品取自新疆维吾尔自治区南疆地区5个县共计4种土类,分别为灌淤土、盐土、潮土、水稻土。所有土样采集深度为表层0~20 cm。采用网格采样法,以样区中心为起点,沿东、南、西、北四个方向采样,每个方向的样点分布于一条直线上,样点之间的距离为50 m左右,每个采集的土样重量约为2 kg。土样带回实验室,去除砾石及动植物残骸等杂质,自然风干,风干后的样品经研磨后过2 mm筛,将处理过的土样按其名称分别装在自封袋中,用于后续盐分测定和光谱测试。

1.3 样品分析

土壤样品的盐分含量采用残渣烘干法进行测定^[18]。表1为供试土壤样品盐分含量的描述统计特征,4种土类按照盐分含量的平均值从低到高排列,可以看出土样的含盐量变化范围较大,在0.325~14.94 g kg⁻¹之间。按照变异系数的等级划分:CV<10%为弱变异性;10%≤CV<100%为中等变异性;CV≥100%为强变异性^[19]。从表1可看出4种土类变异系数均在60%以上,均达到中等变异性,其中潮土变异系数大于100%,属于强变异性,这对于模型的构建是有利的,建立的模型也更具普适性。

1.4 光谱测试

供试土样采用美国FieldSpec Pro FR型光谱仪(美国 Analytical Spectral Devices公司)进行光谱测试,其波长范围为350~2 500 nm,光谱分辨率在350~1 000 nm为3 nm,在1 000~2 500 nm为10 nm。采样间隔为1.4 nm(350~1 000 nm)和2 nm(1 000~2 500 nm)。光谱测量在一个能控制光照条件的暗室内进行。光源是功率为50 W的卤素灯,距土壤表面70 cm。土壤样本分别放置在内部全部涂黑且直径和深度分别为10 cm、1.5 cm

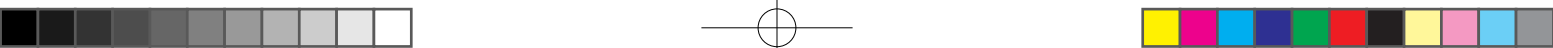


表1 不同类型土壤盐分含量的描述统计特征

Table 1 Descriptive statistics of soil salt contents in soils different in type							
土壤类型 Soil types	采样点 Locations	样本数 Numbers	最小值 Min (g kg ⁻¹)	最大值 Max (g kg ⁻¹)	平均值 Mean (g kg ⁻¹)	标准差 SD (g kg ⁻¹)	变异系数 CV (%)
灌淤土Anthropogenic- alluvial soil	和田/阿瓦提/拜城 Hetian/Awati/Baicheng	135	0.325	6.463	1.306	1.092	80.03
盐土Saline soil	新和Xinhe	47	0.761	11.80	2.151	1.945	90.43
潮土Fluvo-aquic soil	拜城Baicheng	30	0.913	14.94	2.705	3.065	113.3
水稻土Paddy soil	温宿/和田/拜城 Wensu/Hetian/Baicheng	127	0.988	10.55	3.628	4.930	61.20
合计Total	/	339	0.325	14.94	2.485	2.184	87.87

的培养皿内，采用直尺将土壤表面刮平。采用25°视场角的探头，其位于土壤表面垂直上方15 cm处。探头接受土壤光谱为直径6.7 cm的圆形区域，远小于培养皿面积，从而可以保证探头接收的均为土壤的反射光谱。测试之前先进行白板校正，每个土样采集10条光谱曲线，算术平均后得到该土样的实际反射光谱数据。

1.5 光谱数据预处理

每个土壤样品的光谱曲线去除噪声较大的边缘波段350 ~ 399 nm和2 401 ~ 2 500 nm，保留400 ~ 2 400 nm的反射光谱数据用于后续研究。借鉴近几年有关学者的研究方法^[20-23]，对去噪后的数据尝试进行Savitzky-Golay平滑加一阶导或二阶导、多元散射校正MSC（multiplicative scatter correction）、标准正态变量变换SNV（standard normal variate）、归一化（normalization）、标准化（standardization）、均值中心化（mean centering）、开二次方根 $\sqrt{R}$ 、对数 $\log R$ 、对数的倒数 $1/\log R$ 、倒数的一阶导 $(1/R)'$ 、倒数的二阶导 $(1/R)''$ 、对数的一阶导 $(\log R)'$ 、线性基线校正LBC（linear baseline correction）、卷积平滑SGS（savitzky-golay smoothing，3点平滑）、连续统去除CR（continuum removal）等预处理方法。通过后期光谱分类效果和预测效果的对比，本研究选用范围归一化RN（range normalization）数据预处理方法。为压缩高光谱数据，采用主成分分析方法（PCA）对预处理后的光谱进行降维。

Kennard-Stone（K-S）方法^[24-25]基于光谱变量间的欧氏距离，在特征空间中均匀地选取光谱差异较大的样本。该方法的特点是能保证校正集中样

本按照空间距离均匀分布，但需要进行数据自动调整。与传统的随机挑选校正集进行建模相比，K-S方法挑选具有代表性的校正集，不但可以减少建模的工作量，而且直接影响所建模型的适应性和准确性。

1.6 建模方法及评价指标

光谱分类是将具有相似光谱反射率和曲线特征的土壤分为一类，常用的方法包括系统聚类分析法、*k*均值聚类法、模糊聚类法和自组织（Kohonen）神经网络等方法^[26]。在进行土壤光谱分类时，核心问题是怎样找出最佳的分类数目，目前很多聚类分析方法均需要人为设定阈值或本领域的专家参与分析判别来确定，还不能给出最佳类别数目的定量判别指标。本研究引入模糊*k*-均值聚类是一种常用的非监督聚类方法，它的基本思想是将一个数据集分成*k*个类别，寻找目标函数的迭代最小化，其优势是能给出最佳分类数的分析指标^[20, 27-28]。模糊*k*-均值具体算法参见文献^[29]。

本研究借助模糊性能指数FPI、改进分类熵MPE和聚类独立性指数S这三个指标来确定最佳的光谱分类数^[28]，FPI值在0 ~ 1之间变化，若该值接近0时，表示聚类时共用数据较少，类的划分明显；反之则类的划分不明显。三个指标均是值越小越好，FPI越小表示聚类中的共用光谱数据越少，MPE越小表示分解量越大，S越小表示分类精度越高。若FPI和MPE的值在某个类别时均最小，表示该类是最佳分类数^[30]。因此本研究通过对不同分类数目进行聚类分析，然后根据这三个指标来确定最佳分类数目。

预测模型采用外部验证对其稳定性和预测精度进行评价,评价指标为建模决定系数(R_c^2)、建模均方根误差($RMSE_c$)、预测决定系数(R_p^2)、预测均方根误差($RMSE_p$)、相对分析误差(RPD)和RPIQ(RPIQ为IQ与RMSE的比值,其中IQ为样本观测值第三四分位数Q3与第一四分位数Q1的差)为精度评价指标。 R_c^2 、RPD及RPIQ越大, $RMSE_p$ 越小,说明预测效果越好^[31]。针对RPD而言,当 $1.5 < RPD < 2.0$ 时表明模型只能对样品高含量与低含量进行粗略估测,当 $2.0 < RPD < 2.5$ 时表明模型具有较好的预测能力,当 $2.5 < RPD < 3.0$ 时模型具有很好的预测能力,当 $RPD > 3.0$ 时模型具有极好的预测能力^[32]。Bellon-Maurel等^[33]指出土壤理化数据一般存在非正态分布,对于非正态土壤理化数据而言,RPIQ指数较RPD评价更为客观。

研究所用到的模糊 k -均值聚类、光谱预处理、主成分分析、PLSR建模和验证等分别在FuzzyME

3.5、The Unscrambler X10.3软件、SPSS 18.0和Matlab 2013a软件中完成,CR在ENVI 5.0中实现。

2 结果与讨论

2.1 不同类型土壤高光谱特征与分类

所有的土壤样本按照土类进行划分,求得每种土类的平均反射率光谱曲线。由图1a可见不同类型土壤光谱曲线存在差异,其中灌淤土、潮土和水稻土这三个土类的光谱,其反射强度、曲线变化的斜率和特征吸收谷的强度等基本相似,均值光谱曲线在400~780 nm部分呈陡坎型。相较于其他土类的光谱曲线在整个波段内有重合或相互缠绕,盐土光谱曲线除了在400~510 nm波段几乎与潮土重合外,其后的波段反射率均明显低于其他土类。发现在1 650~2 200 nm波段内随着盐分含量的增加,土壤反射率有下降的趋势,这是由于该研究针对的是不同类型土壤。

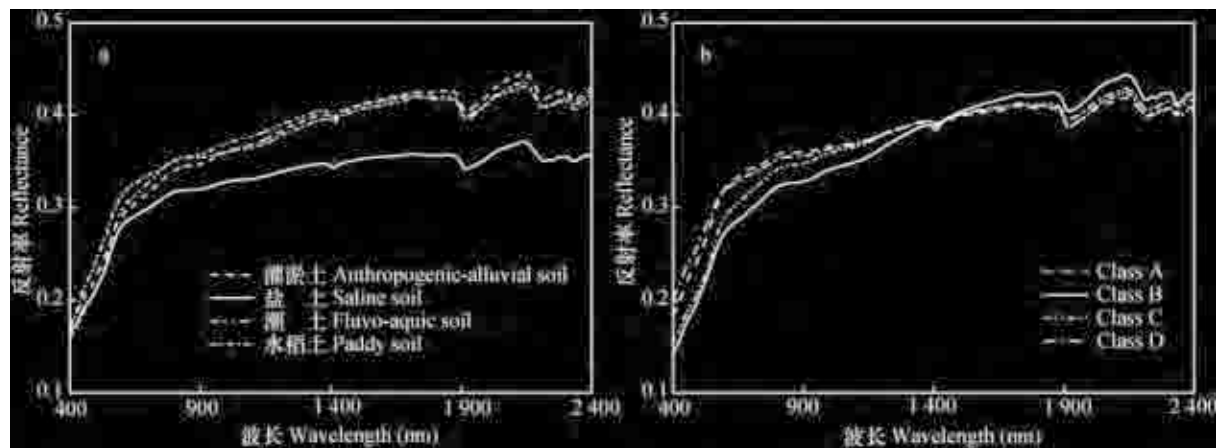


图1 不同类型土壤的反射率均值曲线(a)和土壤光谱分类后的反射率均值曲线(b)比较

Fig. 1 Comparison between different types of soils in reflectance (a) and in mean reflectance curve (b) after classification of soil spectra

鉴于土壤高光谱数据所包含的信息丰富,因此数据量比较大,各波段之间又存在多重共线性问题,针对预处理后的全部339个土样,在此引入模糊 k -均值聚类方法进行土壤光谱分类并找出最佳分类数目。首先采用主成分分析方法进行数据降维处理,得到主成分1(PCA1, 98.81%)、主成分2(PCA2, 0.63%)和主成分3(PCA3, 0.41%),累积贡献率达99.85%。然后取前三个主成分数据作为模糊 k -均值聚类分析的输入数据。取最大迭代次数为300,收敛阈值为0.000 1,模糊加权指数为1.5^[30]。为了找出最佳分类数目,将全部样本

划分为2~9个类别。计算每一类的模糊性能指数、改进分类熵和聚类独立性指数,三者均是越小越好。图2是全部土壤样本按照反射光谱特征划分为2~9类时得到的FPI、MPE和S值。从图中可看出,当分类数目为4时,FPI、MPE和S值均为最小,因此本研究最佳的分类数目为4。图1b为将全部土样光谱分为4类(记为Class A、Class B、Class C和Class D)时每类土样的均值光谱曲线,可以看出4类光谱曲线的反射强度、曲线变化的斜率和特征吸收谷强度基本相似,各类均值曲线总体差异不大,但层次较为分明,Class B在400~1 443 nm波段内

较其他三类光谱曲线低,而在1 444~2 400 nm波段内又明显高于之前。其中Class B由73个土样组成,全部来自和田县的灌淤土(50个)和水稻土(23个)两个土类共73个。Class A由95个土样组成,主要由新和县的盐土(44个)和温宿县的水稻土(48个)两个土类共92个,约占该类型所有土样的97%。Class C由83个土样组成,主要由拜城县的灌淤土(25个)、潮土(30个)和温宿县的水稻土(16个)三个土类共71个,约占该类型土样的86%。Class D由88个土样组成,主要由阿瓦提县的灌淤土(60个)和温宿县的水稻土(27个)两个土类共87个,约占该类型土样的99%。从光谱分类结果与土类之间相互对比可以发现,土壤地带性对土壤光谱分类有着主要控制作用,但也不是绝对的,存在混分现象。例如从Class B和Class C这两个类别来看,前者土样全部来自和田县,而后者中拜城县的土样在该类型中也占有较高的比例,达到66%。这一结论也验证了史舟等^[20]针对有机质进行土壤光谱分类的研究合理性。

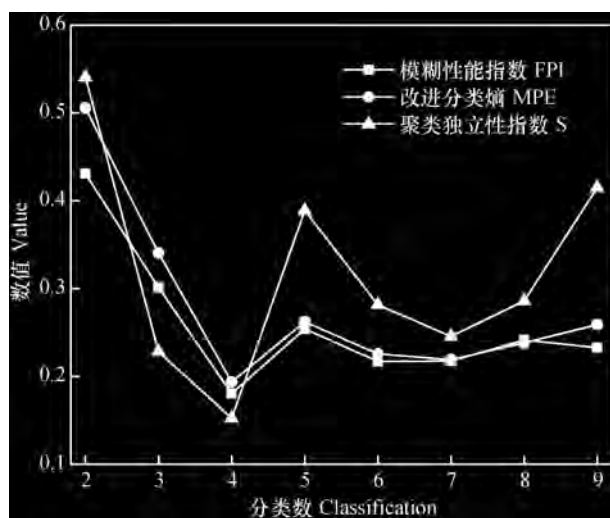


图2 不同类别对应的FPI、MPE和S值

Fig. 2 Calculated FPI, MPE and S value corresponding to various categories of soil spectra

2.2 不同类型土壤盐分高光谱预测

据1.4所述将原始光谱数据进行范围归一化预处理,之后按照土壤类型将变换后的339个土样数据分为4组,且每组按照2:1的比例采用K-S方法划分建模集和预测集,然后将预处理后的每组建模集数据作为输入量,利用PLSR方法分别建立土

壤盐分高光谱预测模型,其预测结果的评价指标见表2。从建模集来看,建模决定系数 R_c^2 均大于0.85,建模均方根误差 $RMSE_c$ 均低于0.62,其中盐土所建立的模型最优, R_c^2 和 $RMSE_c$ 分别为0.998、0.088;从预测集来看,盐土未表现出较优的预测精度,预测精度反而低于另外三类土壤,RPIQ仅为1.106,RPD也低于1.5,表明模型不能对样本进行预测。预测精度较好的是灌淤土所建立的模型, R_p^2 、 $RMSE_p$ 、RPD和RPIQ分别为0.715、0.454、1.894和2.006,但该模型只能对土壤盐分进行粗略估测。针对同一种土壤类型均建立各自的盐分高光谱预测模型,这些土样来自不同地区,由于成土条件和过程等不同,在进行预测时模型的预测精度稳定性并不能保证。

2.3 基于模糊k-均值聚类的土壤盐分高光谱预测

从前述结果可见,依据土壤系统分类各自建立的土壤盐分高光谱模型预测效果并不佳。因此,本研究基于模糊k-均值聚类结果将全部土样光谱数据分成四类,此外在分类前将原始光谱数据进行范围归一化预处理。每类按照2:1的比例采用K-S方法划分建模集和预测集,然后将预处理后的每个类别建模集数据作为输入量,采用PLSR方法分别建立各自类别的盐分高光谱预测模型,最后将未分类的混合土类全局建模(原始光谱、范围归一化(RN))预测结果进行对比(见表3)。

由表3可知,原始光谱-混合土样的全局建模预测结果中, R_p^2 、RPD和RPIQ分别为0.664、1.733、1.461,进行范围归一化预处理后的建模预测精度得到一定的改善, R_p^2 、RPD和RPIQ分别为0.698、1.827、1.924。但是这些全局建模方法RPD均小于2,表明模型只能对土壤盐分高含量和低含量进行粗略估测。因此,本研究采用范围归一化预处理后的数据进行光谱分类建模。由表3可知,光谱分类后PLSR建模的预测精度得到显著提高,总体来看预测的 R_p^2 、RPD和RPIQ分别提高至0.818、2.356、2.422。四个分类所建模型的RPD均大于2.0,其中Class A和Class C这两类所建模型的RPD均高于2.5,表明这两个模型具有很好的预测能力。因此,对于将来在大尺度下新采集的土壤样本,推荐先采用光谱分类,然后选择相应类型的光谱预测模型,这样既可以简化操作流程,又可以显著提高预测精度。



表2 不同类型土壤盐分的PLSR预测结果

Table 2 Soil salt contents in soils different in type predicted with PLSR								
土壤类型 Soil types	建模集Calibration			预测集Prediction			RPD	RPIQ
	样本数 Number	R_c^2	RMSE _c	样本数 Number	R_p^2	RMSE _p		
灌淤土Anthropogenic-alluvial soil	90	0.899	0.344	45	0.715	0.454	1.894	2.006
盐土Saline soil	32	0.998	0.088	15	0.491	0.807	1.451	1.106
潮土Fluvo-aquic soil	19	0.982	0.492	11	0.627	0.797	1.717	1.908
水稻土Paddy soil	85	0.873	0.628	42	0.700	0.943	1.825	2.567
合计Total	226	0.688	1.002	113	0.713	1.295	1.873	1.925

注： R_c^2 为建模决定系数； R_p^2 为预测决定系数；RMSE_c为建模均方根误差；RMSE_p为预测均方根误差；RPD为相对分析误差；RPIQ为IQ与RMSE_p的比值，其中IQ为样本观测值第三四分位数Q3与第一四分位数Q1的差。下同Note: R_c^2 stands for modeling determination coefficient; and R_p^2 for prediction determination coefficient; RMSE_c for root mean square error of the modeling; RMSE_p for root mean square error of the prediction; RPD for relative percent deviation; RPIQ for ratio of IQ and RMSEP, (IQ (=Q3-Q1)); Q1 for the first quartile; and Q3 for the third quartile. The same below

表3 光谱全局建模与分类建模的土壤盐分高光谱预测精度对比

Table 3 Comparison of the global spectral model with the category-based spectral models in prediction accuracy using of soil salinity hyperspectra

类别 Classification	建模集Calibration			预测集Prediction			RPD	RPIQ
	样本数 Number	R_c^2	RMSE _c	样本数 Number	R_p^2	RMSE _p		
原始光谱-混合土类 Original-Mixed	226	0.708	0.989	113	0.664	1.219	1.733	1.461
RN-混合土类 RN-Mixed	226	0.716	0.977	113	0.698	1.014	1.827	1.924
Class A	65	0.885	0.578	30	0.867	0.905	2.785	2.808
Class B	49	0.957	0.607	24	0.823	1.320	2.445	3.669
Class C	55	0.995	0.103	28	0.874	0.575	2.869	3.448
Class D	59	0.954	0.251	29	0.799	0.597	2.265	2.485
合计Total	228	0.802	0.921	111	0.818	1.132	2.356	2.422

3 结 论

本研究通过对来自南疆地区的温宿县、拜城县、阿瓦提县、新和县、和田县4种土类339个土样的室内高光谱反射数据的分析表明，不同类型土壤的反射光谱均值曲线存在差异性，但灌淤土、潮土和水稻土的反射强度、曲线变化的斜率和特征吸收谷的强度等基本相似，说明它们还有很多共性问题，但土壤有机质、氧化铁、盐类等组分不同均会影响光谱曲线的走势。从光谱分类结果来看，同一

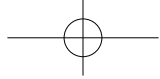
土壤类型被分到不同类别，不同土壤类型组合成新的类别，从而表现出来同物异谱与异物同谱的特征。所以，土壤反射光谱只能说明土壤部分理化性质差异，并不能很好地揭示不同土类之间的差异。因此，土壤反射光谱只可以作为土壤分类的一个间接指标。研究表明模糊k-均值聚类方法将全部土样按照反射光谱特征分为四种光谱类型，结合PLSR方法进行分类建模，可以显著提高模型的预测精度和稳定性，改善了依靠人为判别或依据传统土壤分类的不足，同时也为土壤养分等理化性质的光谱建

模及预测提供了新的思路和方法。

参考文献

- [1] Zhou S W, Zhang G Y, Zhang X N. Exchange reaction between selenite and hydroxyl ion of variable charge soil surfaces: I. Electrolyte species and pH effects. *Pedosphere*, 2003, 13 (3) : 227—232
- [2] 祖皮艳木·买买提, 海米提·依米提, 吕云海. 于田绿洲典型区土壤盐分及盐渍土的空间分布格局. *土壤通报*, 2013, 44 (6) : 1314—1320
Mamat Z, Yimit H, Lü Y H, et al. Spatial distributing pattern of salinized soils and their salinity in typical area of Yutian oasis (In Chinese). *Chinese Journal of Soil Science*, 2013, 44 (6) : 1314—1320
- [3] 俞仁培. 对盐渍土资源开发利用的思考. *土壤通报*, 2001, 32 (S1) : 138—140
Yu R P. Ideas about exploitation and utilization of saline soil resources (In Chinese). *Chinese Journal of Soil Science*, 2001, 32 (S1) : 138—140
- [4] 彭杰, 迟春明, 向红英, 等. 基于连续统去除法的土壤盐分含量反演研究. *土壤学报*, 2014, 51 (3) : 459—469
Peng J, Chi C M, Xiang H Y, et al. Inversion of soil salt content based on continuum-removal method (In Chinese). *Acta Pedologica Sinica*, 2014, 51 (3) : 459—469
- [5] 郭全恩. 土壤盐分离子迁移及其分异规律对环境因素的响应机制. 陕西杨凌: 西北农林科技大学, 2010
Guo Q E. The response mechanism of soil salt ions transfer and diversity rule to environment factor (In Chinese). Yangling, Shaanxi: Northwest A&F University, 2010
- [6] 高凤菊. 盐度对不同类型甜高粱品种萌发、生长发育及产量的影响. 山东泰安: 山东农业大学, 2011
Gao F J. Effects of salinity on germination, growth and development, and yield of sweet sorghum (In Chinese). Taian, Shandong: Shandong Agricultural University, 2011
- [7] 田长彦, 周宏飞, 刘国庆. 21世纪新疆土壤盐渍化调控与农业持续发展研究建议. *干旱区地理*, 2000, 23 (2) : 177—181
Tian C Y, Zhou H F, Liu G Q. The proposal on control of soil salinizing and agricultural sustaining development in 21's century in Xinjiang (In Chinese). *Arid Land Geography*, 2000, 23 (2) : 177—181
- [8] 史舟, 梁宗正, 杨媛媛, 等. 农业遥感研究现状与展望. *农业机械学报*, 2015, 46 (2) : 247—260
Shi Z, Liang Z Z, Yang Y Y, et al. Status and prospect of agricultural remote sensing (In Chinese). *Transactions of the Chinese Society for Agricultural Machinery*, 2015, 46 (2) : 247—260
- [9] 李建国, 濮励杰, 朱明, 等. 土壤盐渍化研究现状及未来研究热点. *地理学报*, 2012, 67 (9) : 1233—1245
Li J G, Pu L J, Zhu M, et al. The present situation and hot issues in the salt-affected soil research (In Chinese). *Acta Geographica Sinica*, 2012, 67 (9) : 1233—1245
- [10] Gogé F, Joffre R, Jolivet C, et al. Optimization criteria in sample selection step of local regression for quantitative analysis of large soil nirs database. *Chemometrics and Intelligent Laboratory Systems*, 2011, 110 (1) : 168—176
- [11] Allbed A, Kumar L, Aldakheel Y Y. Assessing soil salinity using soil salinity and vegetation indices derived from ikonos high-spatial resolution imageries: Applications in a date palm dominated region. *Geoderma*, 2014, 230/231: 1—8
- [12] Weng Y L, Gong P, Zhu Z L. Reflectance spectroscopy for the assessment of soil salt content in soils of the Yellow River delta of China. *International Journal of Remote Sensing*, 2008, 29 (19) : 5511—5531
- [13] 屈永华, 段小亮, 高鸿永, 等. 内蒙古河套灌区土壤盐分光谱定量分析研究. *光谱学与光谱分析*, 2009, 29 (5) : 1362—1366
Qu Y H, Duan X L, Gao H Y, et al. Quantitative retrieval of soil salinity using hyperspectral data in the region of Inner Mongolia Hetao irrigation district (In Chinese). *Spectroscopy and Spectral Analysis*, 2009, 29 (5) : 1362—1366
- [14] 王静, 刘湘南, 黄方, 等. 基于ANN技术和高光谱遥感的盐渍土盐分预测. *农业工程学报*, 2009, 25 (12) : 161—166
Wang J, Liu X N, Huang F, et al. Salinity forecasting of saline soil based on ANN and hyperspectral remote sensing (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 2009, 25 (12) : 161—166
- [15] 刘焕军, 张柏, 王宗明, 等. 基于反射光谱特征的土壤盐碱化评价. *红外与毫米波学报*, 2008, 27 (2) : 138—142
Liu H J, Zhang B, Wang Z M, et al. Soil saline-alkalization evaluation basing on spectral reflectance characteristics (In Chinese). *Journal of Infrared and Millimeter Waves*, 2008, 27 (2) : 138—142
- [16] 纪文君, 史舟, 周清, 等. 几种不同类型土壤的VIS-NIR光谱特性及有机质响应波段. *红外与毫米波学报*, 2012, 31 (3) : 277—282
Ji W J, Shi Z, Zhou Q, et al. VIS-NIR reflectance

- spectroscopy of the organic matter in several types of soils (In Chinese). *Journal of Infrared and Millimeter Waves*, 2012, 31 (3): 277—282
- [17] 王乾龙, 李硕, 卢艳丽, 等. 基于大样本土壤光谱数据库的氮含量反演. *光学学报*, 2014, 34 (9): 308—314
Wang Q L, Li S, Lu Y L, et al. Nitrogen content inversion based on large sample soil spectral library (In Chinese). *Acta Optica Sinica*, 2014, 34 (9): 308—314
- [18] 鲍士旦. 土壤农化分析. 北京: 中国农业出版社, 2000
Bao S D. *Soil and Agrochemical analysis* (In Chinese). Beijing: China Agriculture Press, 2000
- [19] 吕真真, 刘广明, 杨劲松. 新疆玛纳斯河流域土壤盐分特征研究. *土壤学报*, 2013, 50 (2): 289—295
Lü Z Z, Liu G M, Yang J S. Soil salinity characteristics of Manas River Valley in Xinjiang (In Chinese). *Acta Pedologica Sinica*, 2013, 50 (2): 289—295
- [20] 史舟, 王乾龙, 彭杰, 等. 中国主要土壤高光谱反射特性分类与有机质光谱预测模型. *中国科学: 地球科学*, 2014, 44 (5): 978—988
Shi Z, Wang Q L, Peng J, et al. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations (In Chinese). *Science China: Earth Sciences*, 2014, 44 (5): 978—988
- [21] 陈颂超, 冯来磊, 李硕, 等. 基于局部加权回归的土壤全氮含量可见-近红外光谱反演. *土壤学报*, 2015, 52 (2): 312—320
Chen S C, Feng L L, Li S, et al. Vis-NIR spectral inversion for prediction of soil total nitrogen content in laboratory based on locally weighted regression (In Chinese). *Acta Pedologica Sinica*, 2015, 52 (2): 312—320
- [22] 张俊华, 马天成, 贾科利. 典型龟裂碱土土壤光谱特征影响因素研究. *农业工程学报*, 2014, 30 (23): 158—165
Zhang J H, Ma T C, Jia K L. Factors affecting spectral characteristics of typical takyr solonetz (In Chinese). *Transactions of the Chinese Society of Agricultural Engineering*, 2014, 30 (23): 158—165
- [23] 王海江, 张花玲, 任少亭, 等. 基于高光谱反射特性的土壤水盐状况预测模型研究. *农业机械学报*, 2014, 45 (7): 133—138
Wang H J, Zhang H L, Ren S T, et al. Prediction model of soil water-salt based on hyperspectral reflectance characteristics (In Chinese). *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45 (7): 133—138
- [24] Kennard R W, Stone L A. Computer aided design of experiments. *Technometrics*, 1969, 11 (1): 137—148
- [25] 刘伟, 赵众, 袁洪福, 等. 光谱多元分析校正集和验证集样本分布优选方法研究. *光谱学与光谱分析*, 2014, 34 (4): 947—951
Liu W, Zhao Z, Yuan H F, et al. An optimal selection method of samples of calibration set and validation set for spectral multivariate analysis (In Chinese). *Spectroscopy and Spectral Analysis*, 2014, 34 (4): 947—951
- [26] 褚小立. 化学计量学方法与分子光谱分析技术. 北京: 化学工业出版社, 2011
Chu X L. *Molecular spectroscopy analytical technology combined with chemometric and its applications* (In Chinese). Beijing: Chemical Industry Press, 2011
- [27] Gorsevski P V, Gessler P E, Jankowski P. Integrating a fuzzy k -means classification and a bayesian approach for spatial prediction of landslide hazard. *Journal of Geographical Systems*, 2003, 5 (3): 223—251
- [28] Guo Y, Shi Z, Li H Y, et al. Application of digital soil mapping methods for identifying salinity management classes based on a study on coastal central China. *Soil Use and Management*, 2013, 29 (3): 445—456
- [29] Bezdek J C. *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981
- [30] 郭燕, 田延峰, 吴宏海, 等. 基于多源数据和模糊 k -均值方法的农田土壤管理分区研究. *土壤学报*, 2013, 50 (3): 441—447
Guo Y, Tian Y F, Wu H H, et al. Zoning of soil management based on multi-sources data and fuzzy k -means (In Chinese). *Acta Pedologica Sinica*, 2013, 50 (3): 441—447
- [31] 纪文君, 李曦, 李成学, 等. 基于全谱数据挖掘技术的土壤有机质高光谱预测建模研究. *光谱学与光谱分析*, 2012, 32 (9): 2393—2398
Ji W J, Li X, Li C X, et al. Using different data mining algorithms to predict soil organic matter based on visible-near infrared spectroscopy (In Chinese). *Spectroscopy and Spectral Analysis*, 2012, 32 (9): 2393—2398
- [32] Saeys W, Mouazen A M, Ramon H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering*, 2005, 91 (4): 393—402
- [33] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by nir spectroscopy. *Trends in Analytical Chemistry*, 2010, 29 (9): 1073—1081



Prediction on Soil Salt Content Based on Spectral Classification

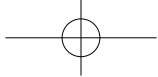
DAI Xijun^{1, 2} PENG Jie^{3†} ZHANG Yanli³ LUO Huaping^{1, 2} XIANG Hongying³

(1 College of Mechanical and Electrical Engineering, Tarim University, Alar, Xinjiang 843300, China)

(2 Key Laboratory of Modern Agricultural Engineering, Tarim University, Alar, Xinjiang 843300, China)

(3 College of Plant Science, Tarim University, Alar, Xinjiang 843300, China)

Abstract 【Objective】 Soil salinization is an issue related to the strategy of sustainable development and improvement of environment quality in arid regions. Soil salinization and irrigation-triggered soil secondary salinization are problems that severely restrict further development of the agriculture in Xinjiang and also the major factors that affect stability of the oasis ecological environment. Therefore, it is of profound significance to sustainable development of the agriculture in Xinjiang and in China as well to manage and ameliorate salinized arable lands and readjust and optimize planting structure in the region, moreover, the information of soil salinization is the fundamental basis for formulation and implementation of these policies. How to acquire the dynamic information of large areas of salinized soil quickly, accurately and inexpensively and thereby formulate rational policies and strategies for management and amelioration of salinized soils are new themes of the study of agricultural science. 【Method】 Hyperspectral remote sensing is a relatively ideal means of monitoring soil. A series of tests and experiments of monitoring soil salinization with near-ground hyperspectra were performed, thus providing a theoretical basis and technical support for application of the technique of hyperspectral remote sensing to monitoring soil salinization. Based on similar soil components generating similar spectral characteristics, soil spectra is to be classified and at the same time effective information to be fully mined by making use of the data and curve characteristics of soil spectral reflectance, which is one of the importance fields for application of spectral analysis. To that end a total of three hundred and thirty-nine topsoil samples (0 ~ 20cm) were collected in Wensu, Baicheng, Awati, Xinhe and Hetian counties of Xinjiang, covering anthropogenic-alluvial soil, saline soil, fluvo-aquic soil, paddy soil. The soil samples were then air dried and ground to pass a 2 nm sieve for determination of salt content with the residue drying method. Hyperspectral reflectivities of the samples were measured separately using the American FieldSpec Pro FR spectrograph (American Analytical Spectral Devices Company), and spectral curves obtained of the samples were pre-treated by erasing the edge bands, i.e. 350 ~ 399 and 2 401 ~ 2 500 nm on the two ends of a curve and keeping the reflection spectral data of the band 400 ~ 2 400 nm for successive research. Clustering of fuzzy k -means is a commonly used unsupervised clustering method, of which the basic idea is to sort a data set into k categories so as to minimize the iteration of target functions and the advantage is the capability of yielding an analysis index for setting an optimal number of categories. 【Result】 By the means of clustering fuzzy k -means, 3 indices, i.e. fuzziness performance index (FPI), modified partition entropy (MPE) and clustering independence index (S) (all being the lower the best) were obtained, and the spectral data of the soil samples were sorted into four categories (before sorting, the original spectral range was normalized as pretreatment). Hyperspectral characteristics of the soils different in soil type before and after the sorting were compared and analyzed, and then the soil samples of each category were divided into modeling dataset and prediction dataset at a ratio of 2 : 1 using the Kennard-Stone algorithm. The preprocessed spectral data of the modeling dataset were used as input, and then a global and category-based salinity prediction models



were built up with the PLSR method. The prediction dataset was used to validate the models and evaluate the models in terms of accuracy and stability. Results demonstrate that after the spectral data were classified with the fuzzy k -means clustering method, the Kennard-Stone algorithm was used to further classify the data for modeling, which was significantly higher than the global and category-based models in precision. The overall prediction model determination coefficient (R_p^2), root mean squares error ($RMSE_p$), relative percent deviation (RPD) and ratio of performance to IQ (RPIQ) of the prediction model was increased from 0.664, 1.219, 1.733 and 1.461 to 0.818, 1.132, 2.356 and 2.422, respectively, among which RPD was increased by up to 23.13% and the PRDs of the all the four category-based models were higher than 2.0. 【Conclusion】 The findings demonstrate that it is feasible to predict soil salt contents quantitatively and quite accurately, and that the study has opened up a new way of thought a new method for using large sample spectral data to establish soil attribute prediction models including that for soil salt content, over a large scale area.

Key words Soil; Hyperspectra; Salt; Fuzzy k-means clustering; Salinization; Kennard-Stone method

(责任编辑: 檀满枝)